Recombinant DNA

A Short Course

James D. Watson John Tooze David T. Kurtz

Cover illustration by Marvin Mattelson.

Library of Congress Cataloging in Publication Data
Watson, James D., 1928–
Recombinant DNA.

Includes bibliographies and index.
1. Recombinant DNA. I. Tooze, John. II. Kurtz, David T. III. Title. IV. Title: Recombinant D.N.A.
[DNLM: 1. DNA, Recombinant. QU 58 W339r]
QH442.W373 1983 574.87'3282 83-9069
ISBN 0-7167-1483-3
ISBN 0-7167-1484-1 (pbk.)

Printed in the United States of America

Scientific American Books is a subsidiary of Scientific American, Inc.

Distributed by W. H. Freeman and Company, 41 Madison Avenue, New York, New York 10010

3 4 5 6 7 8 9 0 KP 1 0 8 9 8 7 6 5 4

peptide chains. Specific nucleotide segments (often called "promoters"; page 98) are recognized by RNA polymerase molecules that start RNA synthesis (Figure 3-8a). After transcription of a functional RNA chain is finished, a second class of signals leads to the termination of RNA synthesis and the detachment of RNA polymerase molecules from their respective DNA templates. How termination is brought about is only now being worked out. RNA chains tend to end with several U residues; just before the site of termination, a nucleotide sequence capable of forming a hairpin loop is usually found (Figure 3-8b). At many termination sites, a specific RNA-chain-terminating protein (in *E. coli,* it is called "rho") plays a role.

In the transcription-initiation step, the two chains of the double helix come apart, with only one of two strands at any start site being copied into its RNA complement. The hybrid DNA–RNA double-helical sections generated as transcription proceeds have only a fleeting existence, with the newly made RNA segments quickly peeling off from the transcription complexes, allowing the just-transcribed sections to reassume their native double-helical states. In bacteria, which have no nucleus, the 5′ ends of nascent mRNA chains attach to ribosomes long before their respective chains have been completely synthesized, and as will be seen (page 48), the processes of transcription and translation can be closely coupled.

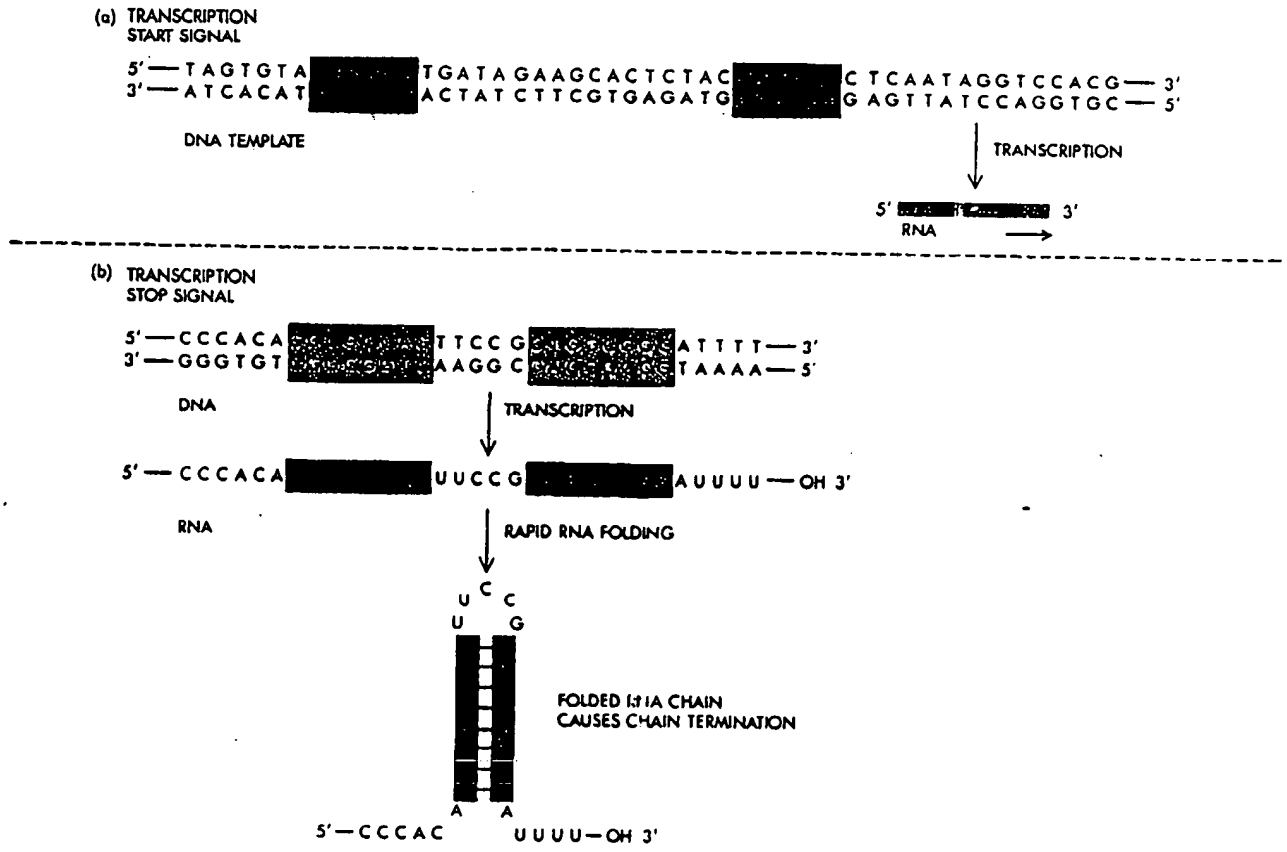


Figure 3-8
(a) RNA polymerase recognizes two blocks of sequences, one at approximately −35 and another at about −10 from the start of transcription. (b) The RNA polymerase stops after transcribing a run of U residues that follows a palindrome. The newly transcribed RNA forms a stem-loop structure that probably constitutes a termination signal for RNA polymerase.

maximum rate of transcription from a particular segment of DNA depends on the sequence of bases in its promoter, and initiation frequency can vary by a factor of at least 1000 (Figure 4-2). Two highly conserved separate sets of nucleotide blocks make up the *E. coli* promoter. One set, consisting of six base pairs (bp), is centered several base pairs upstream of the mRNA start site (at about −10), whereas the second block, of about ten base pairs, is centered some 25 nucleotides further upstream (at about −35). The initial step in *E. coli* transcription is believed to be the recognition and binding of an RNA polymerase molecule to the −35 region. Subsequently, the −10 region is thought to melt (open up) into its component single strands, allowing transcription to begin at the +1 position. Both blocks were initially identified by the existence of point mutations that blocked RNA synthesis. These mutations affect only the synthesis of the mRNA molecule immediately downstream, so promoters are examples of "cis"-acting control elements. In contrast, repressors are not limited in their binding to the DNA molecule that carries their genetic information, so they have been called "trans"-acting control elements.

The operator sequence to which a repressor binds is always close to the promoter for the operon being controlled (Figure 4-1). In some way the binding of a repressor blocks the binding of RNA polymerase to the nearby promoter. The control of a promoter by a repressor is thus an example of negative control. Promoters can also be under the control of positive effectors that increase the rate at which mRNA chains are made. Positive control elements most likely act by helping to open up the two chains of the double helix at the promoter site, thereby facilitating the binding of RNA polymerase.

Constitutive Synthesis of Repressor Molecules

Each repressor is coded by a specific gene. The gene for the lactose repressor lies immediately in front of the operon. In other cases, however, the repressor gene is widely separated from the operon genes on which it acts. The rate at which repressors are made is normally unchanging; such invariant synthesis is known as constitutive synthesis. The exact rate of this constitutive synthesis is a function of the structure of the promoter of the repressor gene. Normally the promoters of repressor genes function at very low rates, leading to the presence of only a few repressor mRNA molecules in the average cell. However, there exist promoter mutants that allow much higher rates of repressor mRNA synthesis and, correspondingly, much

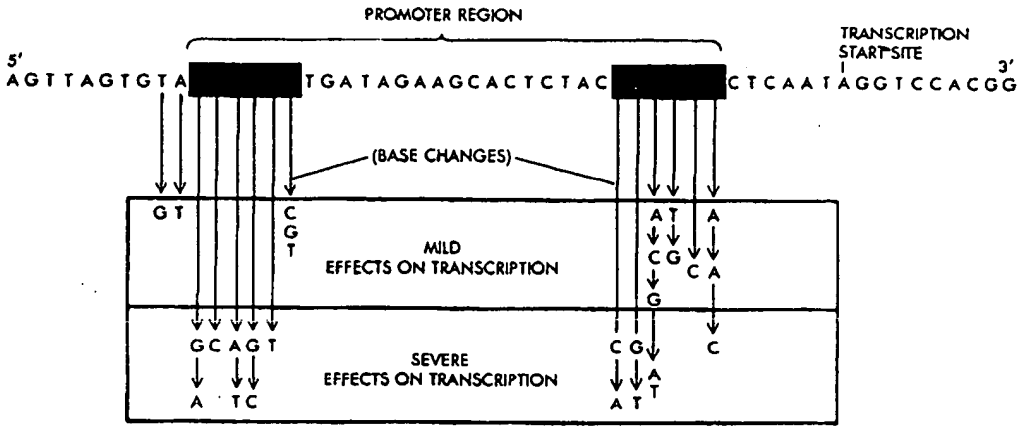
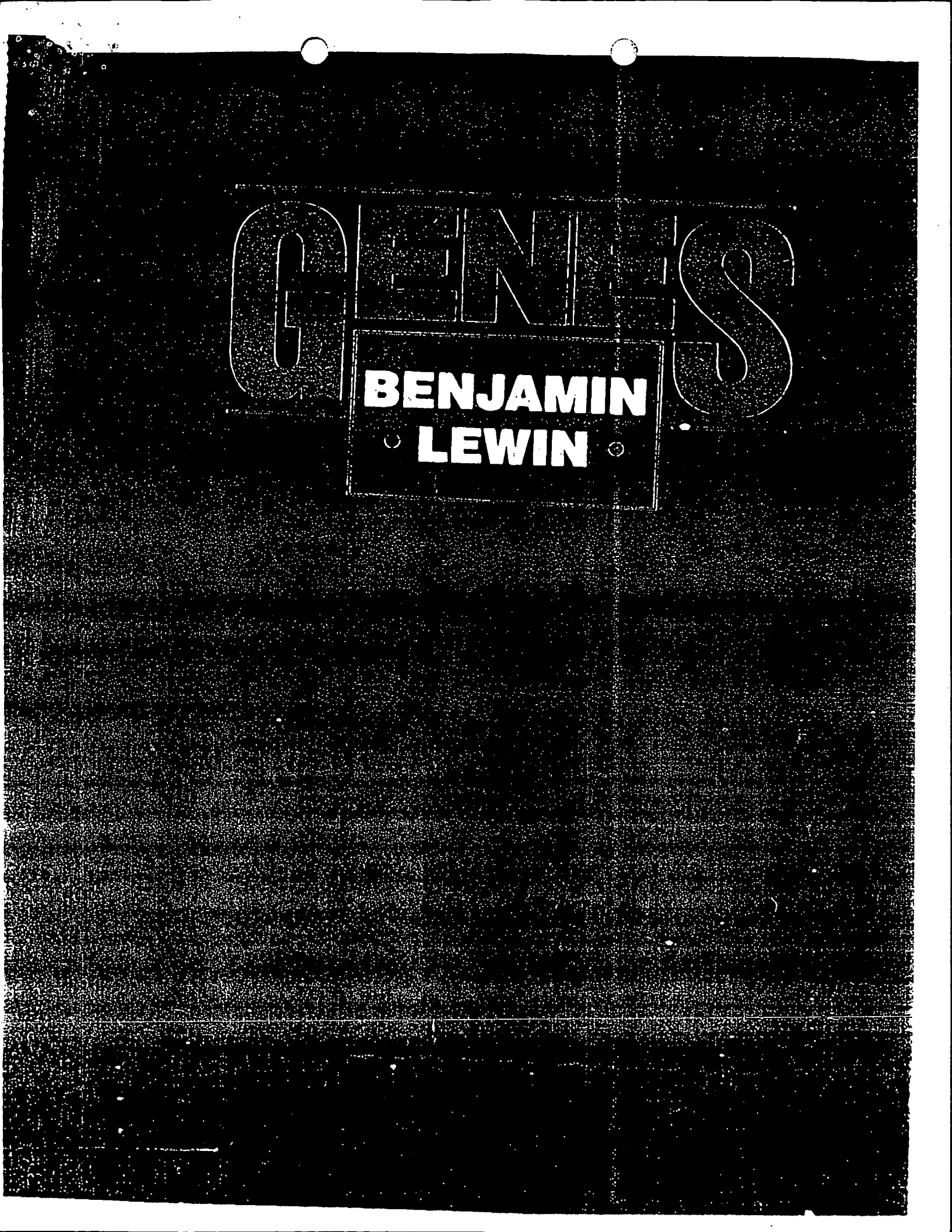


Figure 4-2
Specific DNA sequences are important for efficient transcription of *E. coli* genes by RNA polymerase. The boxed sequences at approximately −35 and −10 are highly conserved in all *E. coli* promoters. Point mutations in these regions have noticeable effects upon transcription efficiency.


GENES
BENJAMIN
LEWIN

1983

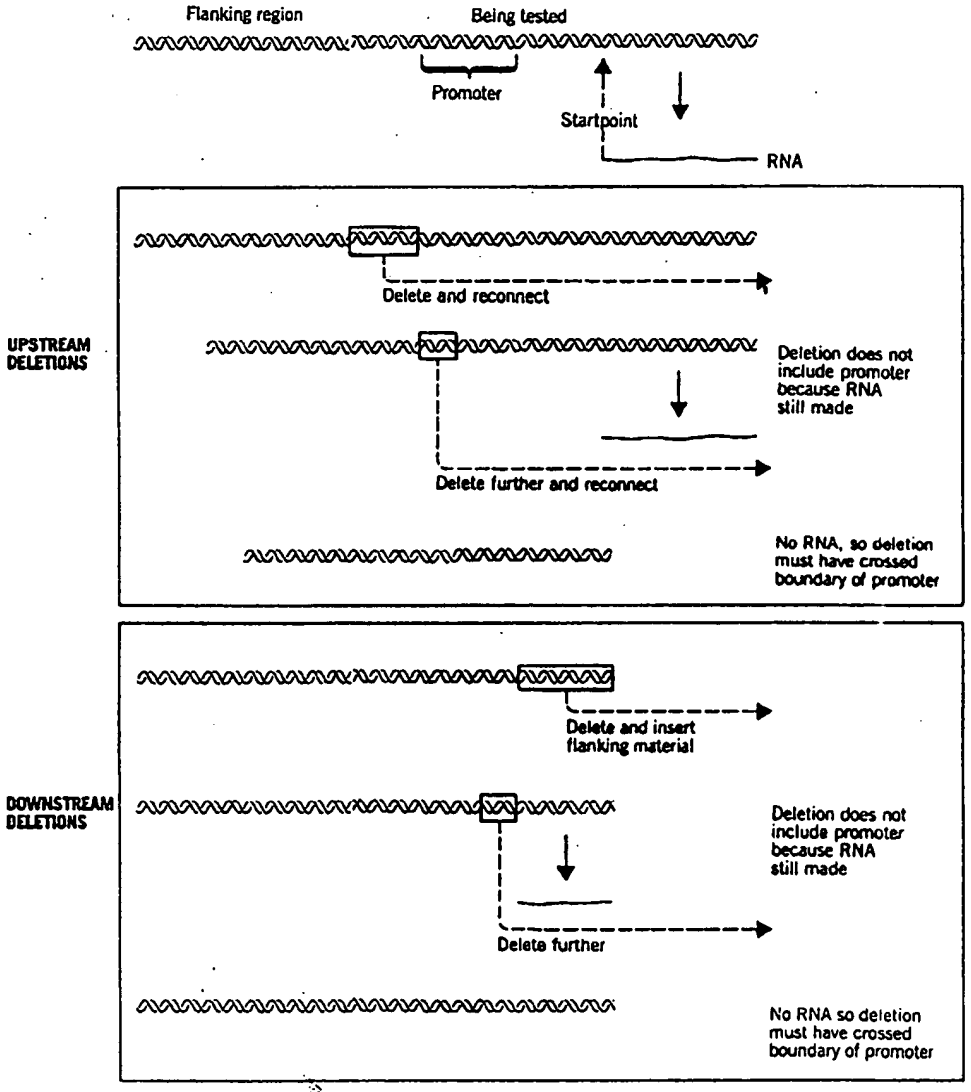


Figure 11.6
Promoter boundaries can be determined by the effects of deletions.
Deletions in the region being tested are made that extend farther into it from either end. When one deletion fails to prevent RNA synthesis but the next stops transcription, the boundary of the promoter must lie between them. (With the downstream mutations, the sequences comprising the transcription unit are altered.)

simply because it happens to be too near the terminus of the DNA fragment. Also, it ensures that the environment is the same in each case, in the form of the identical flanking sequence adjacent to the sequence that is to be tested for promoter function. In the *in vitro* systems, in which termination may not occur properly, the template may be cut at some distance from the promoter (usually about 500 bp downstream), to en-

sure that all polymerases "run off" in the same way, to generate an identifiable transcript.

Once the boundaries of the promoter have been defined, the importance of particular bases within it can be determined by introducing point mutations or other rearrangements in the sequence. As with the bacterial RNA polymerase, these can be characterized as *up* or *down* mutations. Some of these rearrangements may affect only the *rate* of initiation; others may influence the *site* at which initiation occurs, as seen in a change of the startpoint. To be sure that we are dealing with comparable products, in each case it is necessary to characterize the 5′ end of the RNA, as described previously in Figure 11.2.

RNA POLYMERASE II FUNCTIONS ACCURATELY *IN VITRO*

Results with the *in vitro* systems are mostly fairly straightforward. Proceeding from the upstream direction, sequences can be progressively removed without any effect, until reaching a point located somewhere between −45 and −30. The exact limit varies with the system, but is always to the left of the TATA box. Once the limit is transgressed, transcription is reduced twentyfold or more. This is analogous to our concept of the bacterial promoter, which represents a short and well-defined sequence just upstream from the startpoint.

Proceeding from the downstream direction, the boundary of the promoter varies. In two well-characterized cases, it lies at about −10 or −12. This is well upstream from the startpoint, whose sequence must therefore be relatively less important. In such cases, when the startpoint itself has been deleted, initiation occurs at a point in the template that is the *same distance* from the promoter as the original startpoint, except that it may be adjusted by a base or two in order to find a purine (usually an A residue) with which to start. However, the efficiency of initiation may be somewhat reduced by the absence of the usual startpoint. In another case, the essential sequence extends to +6 and so includes the startpoint.

We do not know exactly what features of the region defined by these boundaries are necessary for recognition by RNA polymerase. The apparent failure always to include the startpoint could mean that the geometry of the complex may define where initiation occurs, presumably because the enzyme binds to the upstream site in such a way as to stretch downstream. The TATA box always lies within the region, and its importance is confirmed by the results of introducing two mutations into the (chick) conalbumin gene. Changing the T in the third position to a G provides a strong down mutation. It is not just the change in base-pair composition that is effective, because a change to an A has the same effect (this has simply reversed the orientation of the T-A pair). Thus the exact sequence of the TATA box may be important. The adequacy of this region in the *in vitro* assay is demonstrated by the ability of the sequence from −32 to −12 of the late adenovirus transcription unit to cause transcription to initiate about 30 bp downstream from wherever it is located in a bacterial plasmid.

All of these results were obtained with the same *in vitro* system, obtained from cultured mammalian cells. Is it a demonstration of evolutionary conservation that this is able to function not just with mammalian genes, but also with avian and insect genes? In the case of the insect gene fibroin, a homologous system from the silk gland of the moth also has been developed; and it demands exactly the same boundaries (−29 to +6). However, the silk gland extract transcribes the fibroin genes about three times better than it transcribes adenovirus; whereas the mammalian cell extract functions about three times better with adenovirus DNA, for which it provides a homologous system. The silk gland system also shows a specific increase in the efficiency with which it transcribes the homologous fibroin gene when the sequences upstream from −73 are retained. The existence of this effect only in the homologous system, together with the tendency of each *in vitro* system to work better with its homologous template, suggests that what we may see *in vitro* is part of a basic promoter function that does work across species barriers, but that represents only part of the full promoter function.

A DISCREPANCY BETWEEN THE *IN VIVO* AND *IN VITRO* BOUNDARIES

With *in vivo* or *in oocyte* systems, we see a strong dependence on sequences farther upstream from the

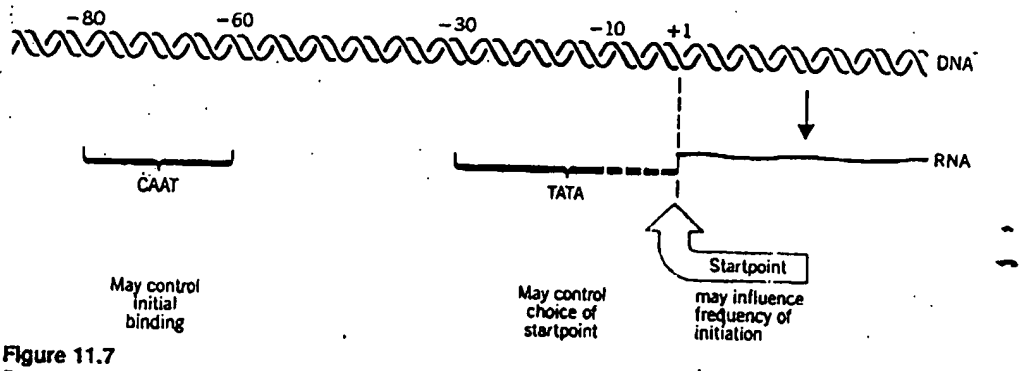


Figure 11.7
Promoters for RNA polymerase II contain separate sequence components with (possibly) different roles. The sequences between the components do not seem to be important.

TATA region. Their exact location varies, but in most cases they are about another 40 bp upstream from the TATA box. Deletion mapping experiments with the TK (thymidine kinase) gene of herpes virus show that a region between −100 and −60 controls the frequency of initiation. In its absence, initiation at the usual startpoint is reduced to 2% of the previous level. If the TATA box is removed, initiation continues to occur, but there is reduced accuracy in selecting the exact startpoint relative to the upstream region. Similarly, in mammalian globin genes, deleting a sequence of roughly 20 to 30 bp, centered about −70, causes a substantial drop in transcription. An upstream sequence also is important in a yeast gene. With a sea urchin histone gene, the sequence lies farther from the startpoint, between −139 and −111. These sequences are not active if the usual region preceding the startpoint is removed, but they increase the frequency of initiation there very considerably.

In vivo, the promoter therefore seems to have two separate components, as illustrated in **Figure 11.7.**

Usually upstream from −60 or thereabouts is an element that has a strong influence on the frequency of initiation, as seen by a drop typically of ten- to fiftyfold in transcription when it is deleted. This element contains the conserved sequence of the CAAT box. Probably it has a major effect on the binding of RNA polymerase. However, the residual transcription that occurs in its absence does initiate at the proper startpoint.

Close to the startpoint is a sequence element that surrounds the conserved TATA box, and whose deletion introduces a more erratic quality into the choice of startpoint, although any overall reduction in transcription is relatively small. The role of this sequence could therefore be to align the RNA polymerase so that it initiates at the proper site. This is consistent with the fact that there are some promoters that naturally lack the TATA sequence, and whose mRNAs may start at more than one point instead of being restricted to the usual single startpoint.

How can a promoter consist of elements separated in this way, and altogether stretching over a greater distance from the startpoint than RNA polymerase could contact on DNA? Two models are illustrated in **Figure 11.8.** One possibility is to think in terms of models that suppose RNA polymerase initially contacts the site farther upstream and then moves to the site nearer the startpoint. Another view requires that we remember that *in vivo* DNA is unlikely to be stretched out in linear fashion; its compact organization could bring into juxtaposition sites that are separated on the duplex molecule (see Figure 29.9). Thus the binding site for RNA polymerase could consist of DNA sequences that are not contiguous, but that are held together in space by proteins that bind DNA. This implies further that the *spacing* (rather than the exact sequence) between these components of the promoter may be important.

In either type of model, we can see that the region farther upstream could be the most important for binding RNA polymerase, but need not necessarily be able to hold the enzyme in a configuration that allows it to

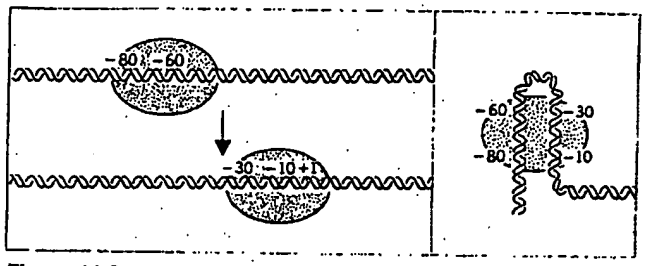


Figure 11.8
One model to reconcile the size of the promoter with the size of RNA polymerase supposes that the enzyme moves; another supposes that the DNA is more compactly organized.

recognize the exact startpoint. This could be the function of the region closer to the startpoint. Thus when the TATA box is deleted, RNA polymerase remains able to bind via the remaining upstream region, but its contacts in the region around the startpoint are less precise, allowing initiation to occur at more than one point. On the other hand, when the sequence farther upstream is absent, the ability of RNA polymerase to bind to the DNA is much reduced; but when it does so, the presence of the TATA box means that initiation is precise.

How are we to explain the discrepancy between the need *in vivo* for the sequence upstream from −60 and its apparent lack of importance *in vitro*? A major responsibility for this effect may lie with the differing efficiencies of transcription in the two circumstances. The *in vitro* system is relatively inefficient; less than 1% of the templates actually are transcribed. We do not know just what proportion of the templates is utilized *in vivo* or in oocyte, but it could well be very much higher. The *in vitro* level could even correspond to the residual expression that we see *in vivo* in the absence of sequences upstream from −60. At all events, the simpler structure of the template *in vitro*, as a DNA molecule not organized in the usual structure, may mean that the manner of recognition by RNA polymerase is different, apparently less efficient, perhaps because it uses only some parts of the promoter.

Even in considering the possibly more complex demands of the *in vivo* system, we have thought of the promoter essentially as an isolated region responsible for binding RNA polymerase. Some further results show that the situation may not be this straightforward. The DNA of the virus SV40 contains two identical s quences of 72 bp each, repeated in tandem abc 200 bp upstream from the usual startpoint of one tra scription unit. They lie in a region that has an unusu nucleoprotein structure (see Chapter 30). Deletic mapping experiments show that the removal of bo of these **72 bp repeats** greatly reduces transcriptic *in vivo*. Transcription is normal so long as one of thes sequences remains. By this type of criterion, we mig argue that this region constitutes an upstream cor ponent of the promoter.

But reconstruction experiments in which the 72 b sequence is removed from the DNA and then is i serted elsewhere show that normal transcription ca be sustained so long as it is present *anywhere* on th DNA molecule. In fact, if a β-globin gene is placed c a DNA molecule that contains the 72 bp sequenc its transcription is increased *in vivo* more than 20(fold, even when the 72 bp sequence is as much a 1400 bp upstream or 3300 bp downstream from th startpoint. And these are simply the limits that hav been tested so far; we have yet to discover at wha distance the 72 bp sequence fails to work.

This sequence is an **enhancer:** it is not part of th promoter, but is able to increase the frequency wit which RNA polymerase initiates. What does it do? Ho can it work over such large distances? One possibili is that the enhancer changes the overall structure c the template—for example, by influencing the DNA protein organization of chromatin, or by changing th density of supercoiling. Another idea is that it migh be responsible for locating the template at a particula place within the cell—for example, attaching it to th nuclear matrix. Still possible, although perhaps les likely, is that it is directly involved in binding RNA poly merase (after which the enzyme would have to mov to the promoter proper).

Several viral genomes have elements that bear a structural resemblance to the SV40 72 bp repeats although there is no homology of sequence. The idea that these elements may fulfill a related function is supported by the demonstration that the analogous component of a retroviral genome can substitute fo the enhancer sequences in SV40 DNA. It may be rel evant that this part of the retroviral DNA is involved ir activating cellular genes.

Enhancer sequences have not at present beer

characterized as natural components of cellular transcription units. The interaction of the viral enhancer with a cellular promoter may vary, because not all promoters are susceptible to enhancement. Thus the α globin gene promoter is not enhanced by the presence of a 72 bp repeat from SV40.

The moral of these results is clear. We need to be careful in defining the components of the promoter. It is not enough just to show that a deletion of a particular sequence reduces transcription; we must also investigate the importance of the *location* of the deleted sequence. So what is a promoter? If we use a working definition that it constitutes a sequence or sequences of DNA that must be in a fixed location relative to the startpoint, the components around the CAAT and TATA boxes are included, but the enhancer is excluded.

From this definition, we see further that we cannot investigate the internal structure of the promoter simply by making deletions within it, because these could inactivate it by changing critical distances between components, even though the deleted sequences are themselves irrelevant. So we need to make small deletions that are replaced by other sequences of the same length. By this approach, we should be able to delineate those sequences in the general vicinity of the startpoint that are needed by RNA polymerase to initiate transcription.

A DOWNSTREAM PROMOTER FOR RNA POLYMERASE III

Until the promoter of the genes coding for 5S RNA in *X. laevis* was identified, an assumption in all attempts to identify promoter sequences was that the promoter would lie upstream from the startpoint. But in the 5S RNA genes, transcribed by RNA polymerase III, the promoter lies well *within* the transcription unit, more than 50 bases downstream from the startpoint. This location explains the lack of any evident consensus sequences in the upstream regions of genes transcribed by RNA polymerase III.

The 5S RNA gene can be transcribed when a plasmid carrying it is used as template for a nuclear extract obtained from *X. laevis* oocytes, the tissue in which this gene usually is expressed. The promoter was located by using plasmids in which deletions extended into the gene from either direction. The 5S RNA product continues to be synthesized when the entire sequence upstream from the gene is removed. This demonstrates that the promoter does not lie outside the gene.

When the deletions continue into the gene, a product very similar in size to the usual 5S RNA continues to be synthesized so long as the deletion ends before about base +55. The product consists of an RNA, the first part of which represents plasmid DNA; the second part represents the remaining segment of the usual 5S RNA gene sequence. But when the deletion extends past position +55, transcription does not occur. This means that the promoter lies downstream from position +55, but causes RNA polymerase III to initiate transcription at a more or less fixed distance away. In the wild-type gene, initiation occurs at a unique startpoint; when the usual startpoint is absent, initiation occurs at the purine base nearest to the position 55 bp upstream from the promoter.

When deletions extend into the gene from its distal end, transcription is unaffected so long as the first 80 bp remains intact. Once the deletion cuts into this region, transcription ceases. This places the downstream boundary of the promoter at about position +80.

So the promoter for 5S RNA transcription lies between positions +55 and +80 within the gene. A fragment containing this region is able to sponsor initiation of any DNA in which it is placed, at a position about 55 bp before its location. How does RNA polymerase initiate transcription upstream from its promoter? The most likely explanation is that the enzyme binds to the promoter, but is large enough simultaneously to contact regions 55 bp away. As with RNA polymerase II, the geometry of binding to the promoter must dictate the position of the startpoint, subject to the reservation that pyrimidines cannot be used for initiation, so the enzyme has some small flexibility in using the nearest purine to initiate. The difference between the enzymes, of course, is that RNA polymerase II reaches forward to the startpoint from its promoter, whereas RNA polymerase III reaches backward (assuming that the promoter is the binding site).

Are other promoters for RNA polymerase III similarly located within their transcription units? The bounda-

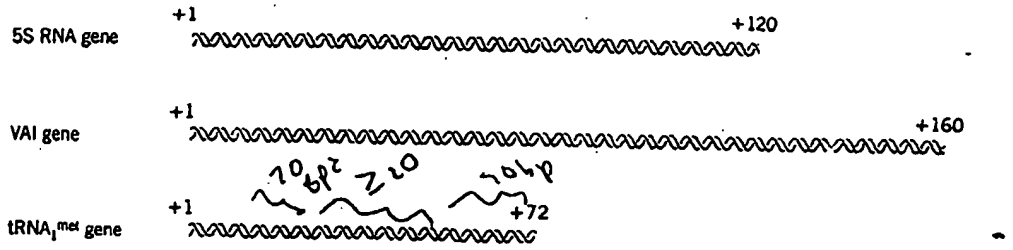


Figure 11.9
Three genes for RNA polymerase III all contain internal promoters (regions in red). The numbers indicate the first and last bases of the gene.

ries of some of them are indicated in **Figure 11.9.** In the VA1 gene of adenovirus, the promoter lies between +9 and +72, as defined by the loss of activity when deletions extend past either boundary. In the $tRNA_1^{Met}$ gene of *X. laevis*, the promoter lies in two separate parts, both within the gene. Deletion mapping shows that both of the sequences from +8 to +30 and from +51 to +72 must be present for transcription. Any deletion between them that reduces their separation prevents initiation, but changes in the intermediate sequence, or insertions that increase the separation of the two components (at least up to 30 bp), have no effect. This promoter therefore consists of two separate regions of about 20 bp each, which must be separated by a minimum distance of 20 bp. Some other tRNA genes have comparable promoters. The sequences of these regions are highly conserved in eucaryotic tRNAs, a fact that had been interpreted solely in terms of tRNA function, but that now also may be attributed to the needs of the promoter.

The internal location of the promoter poses an important question. When a promoter lies outside the transcription unit itself, presumably it is able to evolve freely just to meet the needs of the enzyme. But the sequences needed for initiation of the RNA polymerase III transcription units are constrained also to meet the needs of rather different types of product, including tRNA, 5S RNA, VA RNA, and small nuclear RNAs. How then are they able also to provide whatever features are needed for recognition by RNA polymerase III?

One possible answer is that the enzyme in fact is able to recognize a wide variety of promoter sequences. This would not be entirely unprecedented given the very short consensus sequences found within the promoters recognized by other types of RNA polymerase. An alternative is that the enzyme does not itself recognize the promoters, but acts via an ancillary factor. In this case, there could be a different factor for each type of promoter. In fact, it turns out that RNA polymerase III can transcribe the 5S RNA genes only in the presence of an added factor, a 37,000-dalton protein that binds to the region from +45 to +96. This protein (several copies of which must bind to DNA to cover a region this large) serves a dual purpose: it also binds the 5S RNA product in the oocyte. The ancillary protein does not bind the DNA of the $tRNA_1^{Met}$ gene. We may surmise that other factors of this sort remain to be characterized and will prove to be important in promoter selection.

Although the ability to transcribe these genes is conferred by the internal promoter, the startpoint does have some influence. Changes in the region immediately upstream from the startpoint can alter the efficiency of transcription. In fact, there is one striking case of two tRNA genes whose sequences are identical, but whose flanking upstream regions are different, and which accordingly are transcribed at very different frequencies. Thus the primary responsibility for recognizing the gene lies with the internal promoter, perhaps assisted by ancillary factors; but the responsibility for setting the frequency of initiation lies with the region at the startpoint, perhaps depending on an interaction with RNA polymerase III itself.

FURTHER READING

The interaction of bacterial RNA polymerase with its promoters has been well described by **Gilbert** in *RNA Polymerase* (Losick & Chamberlin, Eds., Cold Spring Harbor Laboratory, New York, 1976, pp. 193–206) and by **Siebenlist, Simpson, & Gilbert** (*Cell* **20,** 269–281, 1980). The analysis of consensus signals has been dealt with by **Rosenberg & Court** (*Ann. Rev. Genet.* **13,** 319–353, 1979). The involvement of supercoiling in promoter function has been reviewed by **Smith** (*Cell* **24,** 599–600, 1981). Some of the techniques and results in characterizing eucaryotic RNA polymerase II promoters have been discussed by **Corden** et al. (*Science* **209,** 1406–1414, 1981); some details of consensus sequences have been dealt with briefly by **Breathnach & Chambon** (*Ann. Rev. Biochem.* **50,** 349–383, 1981). The discrepancy between promoter function *in vivo* and *in vitro* was brought to light by **McKnight** et al. (*Cell* **25,** 385–398, 1981). The transcription of 5S genes has been reviewed by **Korn** (*Nature* **295,** 101–105, 1982). The nature of split promoters for tRNA genes has been analyzed by **Galli, Hofstetter, & Birnstiel** (*Nature* **294** 626–631, 1981) and reviewed by **Hall** et al. (*Cell* **29** 3–5, 1982). A thoughtful analysis of the effects of the SV40 enhancer has been provided by **Banerji, Rusconi, & Schaffner** (*Cell* **27,** 299–308, 1981); the existence of retroviral enhancers has been reported by **Levinson** et al. (*Nature* **295,** 568–572, 1981).

This Page is inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☑ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☑ BLURED OR ILLEGIBLE TEXT OR DRAWING
- ☑ SKEWED/SLANTED IMAGES
- ☐ COLORED OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REPERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.
As rescanning documents *will not* correct images problems checked, please do not report the problems to the IFW Image Problem Mailbox